Plant Gene 7 (2016) 50–61

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plantgene

Genome-wide identification and characterization of the homeodomain-leucine zipper I family of genes in cotton (*Gossypium* spp.)

Jian Zhang [a,b], Qian-Hao Zhu [a,*], Philippe Moncuquet [a], Danny Llewellyn [a], Iain Wilson [a,*]

[a] CSIRO Agriculture, Black Mountain Laboratories, GPO Box 1600, Canberra, ACT 2601, Australia
[b] College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China

ARTICLE INFO

Article history:
Received 13 April 2016
Received in revised form 31 May 2016
Accepted 31 May 2016
Available online 2 June 2016

Keywords:
HD-Zip I transcription factor
Phylogenetic analysis
Gene conversion
Homoeologue
Verticillium dahliae
Cotton

ABSTRACT

Homeodomain-leucine zipper (HD-Zip) transcription factors are unique to the plant kingdom and are classified into four subfamilies, HD-Zip I to IV. This gene family has been extensively investigated in several plant species and many members have been shown to play important roles in plant development and in response to abiotic/biotic stresses. In cotton, several HD-Zip IV genes have been identified and their function investigated, but little is known about the HD-Zip I genes. Here, we performed a genome-wide survey and identified 72, 30 and 34 HD-Zip I genes in *Gossypium hirsutum*, *Gossypium arboreum* and *Gossypium raimondii*, respectively. Almost all *G. arboreum* and *G. raimondii* HD-Zip I genes were retained in allotetraploid *G. hirsutum*, and new HD-Zip I genes were evolved in *G. hirsutum* after polyploidization, probably through tandem and/or segmental duplication. Most HD-Zip I genes were under purifying selection although some could have undergone positive selection. Small indels and nonreciprocal homoeologous recombination (NRHR) events also played a role in shaping the HD-Zip I genes in *G. hirsutum*. Most HD-Zip I genes were preferentially expressed in certain tissues. Differential expression of homoeologues was observed but the differences were generally less than that between different genes. Three HD-Zip I genes were found to have a consistent response in *G. hirsutum* and *G. barbadense* cultivars resistant to *Verticillium dahliae* (*Vd*) following *Vd*-infection. Our results provided a comprehensive view of the cotton HD-Zip I genes and fundamental information for further research towards understanding the role of HD-Zip I genes in cotton.

1. Introduction

The homeodomain (HD) is a conserved 60-amino acid motif, which has a characteristic three-helix structure that is able to bind to specific DNA sequences (Ariel et al., 2007). HD-containing proteins play fundamental roles in a diverse range of developmental processes, from pattern formation to cell type specification, in all eukaryotic organisms (Gehring et al., 1994). Based on the distinguishing features of the HD-encoding sequence, its size and location, as well as other associated domains and gene structures, the HD-containing proteins are classified into six families, including homeodomain-leucine zipper (HD-Zip), Wuschel-related homeobox (WOX), plant homeodomain associated to a finger domain (PHD-finger) or a Bell domain (BELL), Knotted related homeobox (KNOX) and zinc finger-homeodomain (ZF-HD) (Ariel et al., 2007). The HD-Zip transcription factors are unique to the plant kingdom. They contain not only a highly conserved HD, but also a leucine-zipper (LZ) motif immediately downstream of the HD. The HD binds to specific DNA in its target genes while the leucine-zipper motif mediates dimerization (Schena and Davis, 1992). According to the conserved HD-Zip domain, additional conserved motifs, gene structural features and biological functions, the HD-Zip family are further divided into four subfamilies, *i.e.* HD-Zip I to IV (Ariel et al., 2007).

The HD-Zip I proteins contain only the HD domain and an adjacent LZ motif, whereas the HD-Zip II proteins contain five additional conserved amino acids at the C-terminus, known as "CPSCE", and sometimes an additional N-terminal motif (Tron et al., 2002). Members of the HD-Zip I and II subfamilies bind similar pseudo-palindromic *cis* elements, *i.e.* CAAT-NATTG (Sessa et al., 1993; Frank et al., 1998; Johannesson et al., 2001). Both the HD-Zip III and IV subfamily proteins can be distinguished from the HD-Zip I and II subfamilies by the presence of a StAR (steroidogenic acute regulatory protein)-related lipid-transfer (START) domain followed by a START-adjacent domain (Ponting and Aravind, 1999; Schrick et al., 2004); however, the HD-Zip III proteins

Abbreviations: ABA, abscisic acid; dpa, days post anthesis; dpi, days post infection; FPKM, fragments per kilobase of exon per million fragments mapped; HD-Zip, homeodomain-leucine zipper; hpi, hours post infection; *LMI1*, *LATE MERISTEM IDENTITY 1*; NRHR, nonreciprocal homoeologous recombination; *RCO*, *REDUCED COMPLEXITY*; qRT-PCR, quantitative real-time polymerase chain reaction; *Vd*, *Verticillium dahliae*.

* Corresponding authors.
E-mail addresses: zhangjianswau@126.com (J. Zhang), Qianhao.Zhu@csiro.au (Q.-H. Zhu), Philippe.Moncuquet@csiro.au (P. Moncuquet), Danny.Llewellyn@csiro.au (D. Llewellyn), Iain.Wilson@csiro.au (I. Wilson).

http://dx.doi.org/10.1016/j.plgene.2016.05.002
2352-4073/

contain an additional C-terminal MEKHLA domain, while the HD-Zip IV proteins lack this motif (Mukherjee and Bürglin, 2006). The *Arabidopsis thaliana* genome has 17, 9, 5 and 16 genes encoding HD-Zip I, II, III and IV proteins, respectively (Ariel et al., 2007). Similarly, multiple members are present in each HD-Zip subfamily in all plant species investigated, such as maize (Zhao et al., 2011) and soybean (Chen et al., 2014).

The functions of some HD-Zip genes are known. HD-Zip II proteins, for example, have been shown to be involved in responses to illumination conditions, shade avoidance and auxin signaling (Morelli and Ruberti, 2000; Sawa et al., 2002; Rueda et al., 2005; Sessa et al., 2005). HD-Zip III proteins are required for apical meristem and vascular bundle development, embryogenesis, leaf polarity formation and lateral organ initiation (Mattsson et al., 2003; Prigge et al., 2005). HD-Zip IV proteins play crucial roles in anthocyanin accumulation, epidermal cell differentiation, trichome formation, and root and cuticle development (Nakamura et al., 2006; Ariel et al., 2007).

HD-Zip I proteins, on the other hand, have been shown to be important regulators in responses to abiotic/biotic stresses and in the regulation of plant organ growth and development (Ariel et al., 2007). Expression of the HD-Zip I genes has been found to be regulated by a wide range of abiotic stresses, such as drought, extreme temperatures, light conditions, osmosis and phytohormones (Himmelbach et al., 2002; Wang et al., 2003; Olsson et al., 2004; Dezar et al., 2005; Manavella et al., 2006, 2008; Ariel et al., 2010; Zhao et al., 2011; Cabello et al., 2012). For example, *ATHB6*, *ATHB7* and *ATHB12* are either up- or down-regulated by water-deficit conditions and/or externally applied abscisic acid (ABA) in *Arabidopsis*, implying roles in regulating plant responses to dehydration (Lee and Chun, 1998; Söderman et al., 1996, 1999). In *Craterostigma plantagineum*, two HD-Zip I members, *CpHB6* and *CpHB7*, were induced by both drought and ABA, whereas another two members, *CpHB4* and *CpHB5*, were down-regulated by dehydration and were not responsive to ABA treatment (Frank et al., 1998; Deng et al., 2002). *Oshox22*, a HD-Zip I gene in rice, has been shown to affect ABA biosynthesis and regulates drought and salt responses through ABA-mediated signal transduction pathways (Zhang et al., 2012). Similarly, *Zmhdz10*, a HD-Zip I gene in *Zea mays*, can positively regulate drought and salt tolerance through an ABA-dependent signaling pathway (Zhao et al., 2014). *Arabidopsis* plants overexpressing *AtHB13* were resistant to infections with downy mildew (*Hyaloperonospora arabidopsidis*) and green peach aphids (Gao et al., 2014). Regarding the role of HD-Zip I genes in plant development and morphogenesis, the six-rowed spike phenotype in barley (*Hordeum vulgare*) was found to be caused by loss-of-function mutations in *VRS1* (*Six-rowed spike 1*), which suppresses development of the rudimentary lateral spikelets as observed in the two-rowed wild-type barely (Komatsuda et al., 2007). In addition, *Arabidopsis LMI1* (*LATE MERISTEM IDENTITY 1*) and its *Cardamine hirsute* and *Cardamine grandiflora* homologue *RCO* (*REDUCED COMPLEXITY*) have been demonstrated to play a role in leaf serration and leaflet formation (Saddic et al., 2006; Sicard et al., 2014; Vlad et al., 2014). Overexpressing *C. hirsute RCO* in *Arabidopsis* converted simple leaves into complex leaves (Vlad et al., 2014). Our recent work showed that *GhOKRA* (*Gh_D01G2042*), a homologue of *LMI1* and *RCO*, is the gene underlying the okra leaf shape in *Gossypium hirsutum* (Zhu et al., 2016).

Cotton (*Gossypium* spp.) is an important economic crop and a model plant for the study of polyploidy, cell elongation and cell wall synthesis (Paterson et al., 2012). There are at least 50 species in the *Gossypium* genus (Fryxell, 1992). Four species, including *G. hirsutum* (AD_1), *Gossypium barbadense* (AD_2), *Gossypium arboreum* (A_2) and *Gossypium herbaceum* (A_1), are cultivated for their fibers. *G. hirsutum*, also known as Upland cotton, is widely planted in more than 70 countries with China, the United States of America, India and Pakistan being the leading producers, due to its wide adaptability and high production, and yields over 95% of the worldwide cotton fiber. *G. barbadense*, famous for its high fiber quality, produces about 2% of the world's cotton fiber, *G. arboreum* and *G. herbaceum* together contribute about 2% of the world's cotton production (Chen et al., 2007). The allotetraploid cottons (*G. hirsutum* and *G. barbadense*) were derived from a cross between a D-genome species similar to *G. raimondii* (D_5), as a pollen-providing parent, and an A-genome species similar to *G. arboreum* or *G. herbaceum*, as the maternal parent (Wendel and Cronn, 2003).

In addition to *GhOKRA* (Zhu et al., 2016), only a few other HD-Zip genes have been reported and their functions investigated in cotton. *GhHB1*, a HD-Zip I gene, may be involved in response to salt stress and ABA treatment (Ni et al., 2008). *GhHB2/3/4*, encoding HD-Zip II proteins, have been found to be preferentially expressed during the early developmental stages of cotton seedlings and may be involved in phytohormone signaling (Qin et al., 2010). Of the two HD-Zip IV genes identified in *G. arboreum*, *GaHOX1* is predominately expressed in the early fiber developmental stages, while *GaHOX2* is expressed in both fiber and other ovular tissues, including outer and inner integuments (Guan et al., 2008). Of the two HD-Zip IV genes characterized in *G. hirsutum*, *GhHD1* was reported to mainly be involved in trichome development and had only a mild effect on fiber cell development (Walford et al., 2012), whereas *GhHOX3* was reported to have a role in controlling cotton fiber elongation (Shan et al., 2014).

Genome-wide identification of the HD-Zip family genes have been performed in several plant species, including *A. thaliana* (Henriksson et al., 2005), *Oryza sativa* (Agalou et al., 2008), *Zea mays* (Zhao et al., 2011), *Populus trichocarpa* (Hu et al., 2012), *Cucumis sativus* (Liu et al., 2013), *Glycine max* (Chen et al., 2014), *Prunus persica* (Zhang et al., 2014) and *Pyrus betulifolia* (Wang et al., 2015), but no systematic analysis of the HD-Zip genes has been reported in cotton. In this study, we identified HD-Zip I genes in three cotton species (*G. hirsutum*, *G. arboreum* and *G. raimondii*), for which a genome sequence is currently available, with the aims i) to have a genome-wide overview of the HD-Zip I genes in these three cotton species; ii) to understand the evolutionary relationship of the HD-Zip I genes in cotton; and iii) to define the expression profiles of the HD-Zip I genes in different tissues and in response to *Verticillium dahliae* infection in *G. hirsutum*.

2. Materials and methods

2.1. Plant materials and Verticillium dahliae infection

Two *G. hirsutum* varieties (MCU-5 and Siokra 1–4) were used in this study. MCU-5 was used in expression analysis of the HD-Zip I genes in various tissues. MCU-5 (*Verticillium dahliae* resistant) and Siokra 1–4 (*V. dahliae* susceptible) were used in *V. dahliae* infection experiments. Cotton plants were grown in a glasshouse (Canberra, Australia) at 28 ± 2 °C or 22 ± 2 °C (for the *Vd*-infection experiment) with approximately 16 h day and 8 h night regime. Tissues used in gene expression analysis were roots and cotyledons (collected at the cotyledon stage), whole ovules collected at −1, 0, 1, 3 and 5 DPA (days post anthesis) and 15 DPA fiber. To investigate responses of HD-Zip I genes to *V. dahliae* infection, one-true-leaf stage of MCU-5 and Siokra 1–4 seedlings were inoculated with *V. dahliae* as previously described (Zhu et al., 2013). Seedlings treated with water (mock treatment) were used as controls. Leaf samples were collected at 1, 3 and 7 dpi (days post infection) from both *V. dahliae* and mock treated plants. Samples were immediately frozen in liquid nitrogen after collection and stored in −80 °C until RNA was extracted.

2.2. Identification of the HD-Zip I family genes in cotton

Protein sequences of the *Arabidopsis* HD-Zip I genes (17 in total) downloaded from the Arabidopsis Information Resource (TAIR, http://arabidopsis.org/) were used as queries to search against the annotated protein sequences (blastp, E value < 10^{-10}) of *G. raimondii* (Paterson et al., 2012), *G. arboreum* (Li et al., 2014) and *G. hirsutum* (Zhang et al., 2015). After building a list of unique protein sequence hits for each species, we analyzed all sequences to identify HD-Zip I genes by selecting those containing only the homeobox (PF00046) and the homeobox

associated leucine zipper domain (PF02183) but without other conserved domains observed in HD-Zip II to IV proteins, such as the conserved residues "CPSCE" of HD-Zip II and the START domain of HD-Zip III and IV. Protein sequences of the *Arabidopsis* HD-Zip I genes were also used to search against the genome sequences (tblastn, E value < 10^{-10}) of the three cotton species. The tblastn results were compared with those of blastp, which identified two more HD-Zip I genes (*Gh_A05G82568322* and *Gh_D13G7737741*, the number after A05G and D13G represent the coordinate of the translation start site of the predicted protein) in *G. hirsutum*. Annotation of the two newly identified HD-Zip I genes was performed using the online FGENESH program (http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=gfind). No new HD-Zip I genes were identified in both *G. raimondii* and *G. arboreum*.

2.3. Phylogenetic analysis of the HD-Zip I family genes

Protein sequences of all HD-Zip I genes were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and the alignment file was then imported into MEGA6 (http://www.megasoftware.net/) to generate the phylogenetic tree using the neighbor-joining algorithm with the default settings. Subfamilies or clades of the cotton HD-Zip I proteins were classified according to their clustering relationships with the known *Arabidopsis* HD-Zip I proteins.

2.4. Visualization of the chromosomal locations of the HD-Zip I family genes in G. hirsutum

For the annotated HD-Zip I genes, their physical locations were retrieved from the GFF3 file of *G. hirsutum* (Zhang et al., 2015). For the two newly identified HD-Zip I genes, their locations were based on the coordinates identified in the tblastn search using *Arabidopsis* HD-Zip I proteins as queries. The Mapchart 2.2 software was used to visualize the distribution of the HD-Zip I genes on chromosomes of *G. hirsutum*.

2.5. Gene structure analysis

The Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/) was employed to analyze the exon/intron organization of the HD-Zip I genes by alignment of the cDNAs to their corresponding genomic DNA sequences.

2.6. Analysis of dN, dS and dN/dS

Synonymous (*dS*), nonsynonymous (*dN*) substitutions and *dN/dS* ratio were using the codeML module implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) program (http://abacus.gene.ucl.ac.uk/software/paml.html). The parameters used were: runmode - pairwise; seqtype - codons; codonFreq - F3 × 4. The tree file for each HD-Zip I gene used in the analyses was generated by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) using the cDNA sequences of *G. arboreum* (A_2), *G. raimondii* (D_5) and *G. hirsutum* (A_t/D_t). At the end, results for the following pair of HD-Zip I genes, *i.e. G. arboreum* (A_2) *vs G. hirsutum* (A_t) and *G. raimondii* (D_5) *vs G. hirsutum* (D_t), were extracted and shown.

2.7. Gene expression analysis using the publicly available transcriptome data

Two transcriptome datasets [PRJNA248163 (Zhang et al., 2015): RNA-sequencing (RNA-seq) data from various tissues of TM-1 (*G. hirsutum*); PRJNA234454 (Chen et al., 2015): RNA-seq data from 7124 (*G. barbadense*), a *Vd*-resistant cultivar] were downloaded from the SRA database in National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). RNA-seq reads were aligned to the *G. hirsutum* genome sequence (Zhang et al., 2015) using Tophat2 (Kim et al., 2013) with the default setting except for the maximum intron length (changed to 5000 bp) and the maximum number of alignments to be allowed (changed to 1). The latter setting was used to allow the best unique mapping of RNA-seq reads. Quantification of transcript abundance (FPKM: fragments per kilobase of exon per million fragments mapped) was performed using Cufflinks (Trapnell et al., 2012) with the default settings.

2.8. Quantitative real-time polymerase chain reaction (qRT-PCR)

Total RNA was isolated from samples using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. Total RNA was then treated with RQ1 RNase-Free DNase kit to remove possible residual DNA. RNA samples were quantified using a NanoDrop 1000 (Thermo Scientific) and the Qubit-iT RNA Assay Kit (Life Technologies). Two micrograms of DNase-treated total RNA were used in reverse transcription (RT) using the random hexamer primer and SuperScript III reverse transcriptase (Invitrogen). qRT-PCR was performed as previously described (Zhu et al., 2013) except the reference gene used (cotton ubiquitin gene, accession no. EU604080). Gene expression levels in various tissues of MCU-5 were determined based on three biological replicates each with three technical replicates. Gene expression changes in MCU-5 and Siokra 1–4 in response to *V. dahliae* infection were determined based on two biological replicates each with three technical replicates. All qRT-PCR reactions were run on the ABI PRISM™ 7900HT Fast Real-Time PCR system (Applied Biosystems) using the FastStart Universal SYBR Green Master (ROX) (Roche). Relative expression level was calculated using the $2^{-\Delta Ct}$ method.

Primers were designed based on the coding sequences of the HD-Zip I genes of *G. raimondii* using Primer3 (http://simgene.com/Primer3). Of the 34 HD-Zip I genes identified in *G. raimondii*, four were excluded in this study. They are *Gorai.002G244000* (*GhOKRA*) and *Gorai.002G244200*, which have been reported recently in one of our separate studies (Zhu et al., 2016), and *Gorai.004G113700* and *Gorai.008G091100* that are closely related to *Gorai.006G273000* and *Gorai.003G041500*, respectively, but show less similarity to their corresponding D_t subgenome homologues of *G. hirsutum*. Primers designed for the remaining 30 HD-Zip I genes were tested to make sure amplification was of a single product and having a similar PCR efficiency. As a result, five genes did not meet the criteria and thus 25 HD-Zip I genes were finally investigated for their expression profiles in various tissues in MCU-5. For expression changes in response to *V. dahliae* infection, 10 representative HD-Zip I genes were analyzed. All primers used in qRT-PCR analyses are shown in Table S1.

3. Results

3.1. Identification of the HD-Zip I genes in cotton

To identify putative HD-Zip I genes in Upland cotton (*G. hirsutum*), we performed a blastp search against the annotated proteins of *G. hirsutum* using the 17 HD-Zip I protein sequences of *Arabidopsis* (Henriksson et al., 2005) as queries. The cotton protein hits of all *Arabidopsis* HD-Zip I proteins were consolidated to remove redundant hits, which led to a list of 120 unique proteins. Conserved domains in each of the 120 protein were then identified by searches against the Pfam database (http://pfam.xfam.org/). The proteins with only the HD and an immediate downstream LZ domain were considered as putative HD-Zip I types (Fig. S1). In total, 70 HD-Zip I genes (31 and 38 from the A_t and D_t subgenome, respectively, and one without subgenome information) were identified. To identify possible missing HD-Zip I genes, a tblastn search against the entire *G. hirsutum* genome sequence was further performed using the 17 *Arabidopsis* HD-Zip I proteins. Analyses similar to that described above was carried out, which identified two (*Gh_A05G82568322* and *Gh_D13G7737741*) new putative HD-Zip I genes. As a result, 72 HD-Zip I genes (32 and 39 from the A_t and D_t subgenome, respectively, and one without subgenome information)

Table 1
The HD-Zip I family genes in cotton.

	A_t subgenome homoeologue		D_t subgenome homoeologue		A_2 genome homologue		D_5 genome homologue	
Gene name	Gene ID	Coordinates	Gene ID	Coordinates	Gene ID	Coordinates	Gene ID	Coordinates
GhHDZ1	Gh_A01G0439	6934736..6935269	Gh_D01G0446	5263377..5263910	Cotton_A_06253	91883198..91883731	Gorai.002G067100	7797565..7798098
GhHDZ2			Gh_D01G0447	5269938..5270471				
GhHDZ3			Gh_D01G1287	35232845..35233237				
GhHDZ4	Gh_A01G1801	97792773..97793726	Gh_D01G2042	59422879..59468477	Cotton_A_00507	68432568..68433521	Gorai.002G244000	60816695..60817565
GhHDZ5	Gh_A01G1802	97830953..97832191			Cotton_A_00505	68480260..68481502	Gorai.002G244200	60848207..60849955
GhHDZ6	Gh_A02G1237	73523410..73524530	Gh_D03G0374	4933027..4934147	Cotton_A_11191	7279734..7281856	Gorai.003G041500	5032348..5034572
							Gorai.008G091100	21922180..21924344
GhHDZ7	Gh_A02G1563	82209242..82210870	Gh_D03G0161	1217375..1218924	Cotton_A_01632	135751593..135753217	Gorai.003G016600	1149042..1151071
GhHDZ8	Gh_A03G0571	14872225..14873270	Gh_D03G0850	29597891..29598667	Cotton_A_24777	3660749..3662367	Gorai.003G094500	29619205..29620998
GhHDZ9	Gh_A03G0861	49144998..49146521	Gh_D02G2405	15104..16620	Cotton_A_38457	40487157..40488641	Gorai.005G139600	36642833..36645887
GhHDZ10	Gh_A03G1284	89425111..89426325	Gh_D02G1724	58907767..58908978	Cotton_A_19290	30183623..30185596	Gorai.005G189800	55254190..55256571
GhHDZ11	Gh_A04G0563	36981165..36982160			Cotton_A_32808	119846397..119847392	Gorai.009G409600	61505450..61506823
GhHDZ12	Gh_A05G82568322	82568322..82567463	Gh_D04G0457	7391032..7391846	Cotton_A_29012	54386078..54386830	Gorai.012G055400	7419952..7421282
GhHDZ13	Gh_A05G1913	20076808..20077824	Gh_D05G2148	20097210..20098226	Cotton_A_26519	34930892..34932507	Gorai.009G233900	18459497..18461266
GhHDZ14	Gh_A05G2636	40643883..40644900	Gh_D05G2926	34575305..34576320	Cotton_A_30333	53632425..53633446	Gorai.009G323600	31916287..31918347
GhHDZ15	Gh_A06G0266	3312194..3313482	Gh_D06G0289	3272662..3273929	Cotton_A_02815	98555327..98556619	Gorai.010G037800	3524548..3526213
GhHDZ16	Gh_A07G0585	8118408..8119429	Gh_D07G0654	7617440..7618464	Cotton_A_02692	96558487..96556619	Gorai.001G073600	7500077..7501695
GhHDZ17	Gh_A07G1237	28375954..28376752	Gh_D07G1346	21713930..21714727	Cotton_A_32300	18999499..19000297	Gorai.001G152200	21293529..21294639
GhHDZ18	Gh_A07G1966	75673868..75674820	Gh_D07G2184	52588613..52589565	Cotton_A_19429	125340168..125341096	Gorai.001G256300	52989429..52990864
GhHDZ19			Gh_D08G0905	18251112..18251504			Gorai.006G273000	50916674..50917209
							Gorai.004G113700	26205047..26205792
GhHDZ20	Gh_A08G0907	55727791..55728574	Gh_D08G1109	34073960..34074746	Cotton_A_41261	96258442..96259189	Gorai.004G123100	31855133..31856511
GhHDZ21	Gh_A08G0909	56106003..56106783	Gh_D08G1112	34258765..34259545	Cotton_A_39533	107946101..107946881	Gorai.004G123500	32022862..32024223
GhHDZ22			Gh_D08G1853	55687362..55687724				
GhHDZ23	Gh_A09G0195	5710688..5712170	Gh_Sca073981G01	45..618	Cotton_A_27084	85499277..85501013	Gorai.006G022100	5670839..5672935
GhHDZ24			Gh_D09G1239	39421213..39421575				
GhHDZ25			Gh_D10G0770	9157324..9157686				
GhHDZ26	Gh_Sca005979G01	6223..7950	Gh_D10G0847	10657108..10658834	Cotton_A_16168	16837038..16839574	Gorai.011G095900	10561595..10564039
GhHDZ27	Gh_A10G0774	15514968..15516022	Gh_D10G0992	13746841..13747898	Cotton_A_24633	92649729..92651677	Gorai.011G111500	13511064..13513109
GhHDZ28	Gh_A11G0139	1391029..1391759	Gh_D11G0153	1375520..1376254	Cotton_A_02375	21924461..21925191	Gorai.007G016800	1287731..1289216
GhHDZ29	Gh_A11G0410	3804010..3806053	Gh_D11G0473	4067435..4069501	Cotton_A_02021	29581227..29583266	Gorai.007G051100	3606890..3609134
GhHDZ30	Gh_A11G0906	9471489..9472322	Gh_D11G1052	9351818..9352653	Cotton_A_35320	54347480..54349050	Gorai.007G111400	8526352..8528062
GhHDZ31	Gh_A11G1037	11723454..11723750	Gh_D13G1514	47003219..47003515				
GhHDZ32	Gh_A11G1721	26830498..26831896	Gh_D11G1879	22013016..22014350	A_Scaffold212	579663..581060	Gorai.007G206000	21356968..21358666
GhHDZ33	Gh_A11G1855	44657634..44658663	Gh_D11G2140	32234043..32235092	A_Scaffold3987	30900..31957	Gorai.007G235200	31057216..31059291
GhHDZ34			Gh_D11G2639	55114064..55115622			Gorai.007G229700	28335439..28336496
GhHDZ35			Gh_D12G0450	7428749..7429102				
GhHDZ36	Gh_A12G0771	41812683..41813875	Gh_D12G0780	21958355..21959527				
GhHDZ37	Gh_A12G1033	61977282..61978030	Gh_D12G1152	38506621..38507437	Cotton_A_25888	45863864..45864612	Gorai.008G128000	36932582..36933752
GhHDZ38	Gh_A12G2171	84524789..84525856	Gh_D12G2350	56563972..56565024	Cotton_A_13207	15810029..15810847	Gorai.008G258900	54009491..54010823
GhHDZ39	Gh_A12G2639	36458..37524	Gh_D12G1725	49273401..49274472	Cotton_A_25802	100656904..100657971	Gorai.008G190300	47323323..47325215
GhHDZ40	Gh_A13G0639	17340288..17341978	Gh_D13G0756	12170753..12172422	Cotton_A_27957	5897525..5900025	Gorai.013G083600	11713107..11715417
GhHDZ41			Gh_D13G7737741	7737164..7737614				
GhHDZ42	Gh_Sca028226G01 46..959 (uncertain for its subgenome location)							

were identified in *G. hirsutum* (Table 1). Using the same strategy, we identified 30 and 34 HD-Zip I genes in *G. arboreum* (A_2) and *G. raimondii* (D_5), respectively (Table 1). An alignment of the HD and LZ domain sequences the HD-Zip I genes in *G. arboreum* and *G. raimondii* is shown in Fig. S2. In both diploid cotton species, all putative HD-Zip I genes seemed to have been annotated because tblastn searches against the *G. arboreum* and *G. raimondii* genomes did not find new genes of this subfamily.

3.2. Phylogenetic and structural analyses of the cotton HD-Zip I genes

To know the evolutionary relationship amongst the cotton HD-Zip I genes, a phylogenetic tree was constructed using the protein sequences of the HD-Zip I genes identified in *G. hirsutum*, *G. arboreum* and *G. raimondii* as well as the 17 HD-Zip I proteins of *A. thaliana* (Fig. 1). Based on the phylogenetic result and the physical localization of the HD-Zip I genes to 13 pairs of *G. hirsutum* chromosomes, the 72 Upland cotton HD-Zip I genes were designated *GhHDZ1* to *GhHDZ42* (Table 1). Of the 71 genes with a subgenome assigned, 61 were in 30 pairs (*Gh_D01G0446* and *Gh_D01G0447* are tandem identical duplicates homologous to *Gh_A01G0439*), *i.e.* found in the corresponding A_t and D_t homoeologous chromosomes. Two genes (*GhHDZ5* and *GhHDZ11*) were found only in the A_t subgenome, and eight genes (*GhHDZ3*, *GhHDZ19*, *GhHDZ22*, *GhHDZ24*, *GhHDZ25*, *GhHDZ34*, *GhHDZ35* and *GhHDZ41*) were found only in the D_t subgenome (Table 1).

Of the 32 A_t HD-Zip I genes, 30 seemed to be have been inherited from *G. arboreum* and two (*GhHDZ31*/*Gh_A11G1037* and *GhHDZ36*/*Gh_A12G0771*) were newly evolved in *G. hirsutum*. Of the 39 D_t HD-Zip I genes, 30 seemed to have been inherited from *G. raimondii* and nine (*GhHDZ2*/*Gh_D01G0447*, *GhHDZ3*/*Gh_D01G1287*, *GhHDZ22*/*Gh_D08G1853*, *GhHDZ24*/*Gh_D09G1239*, *GhHDZ25*/*Gh_D10G0770*, *GhHDZ31*/*Gh_D13G1514*, *GhHDZ35*/*Gh_D12G0450*, *GhHDZ36*/

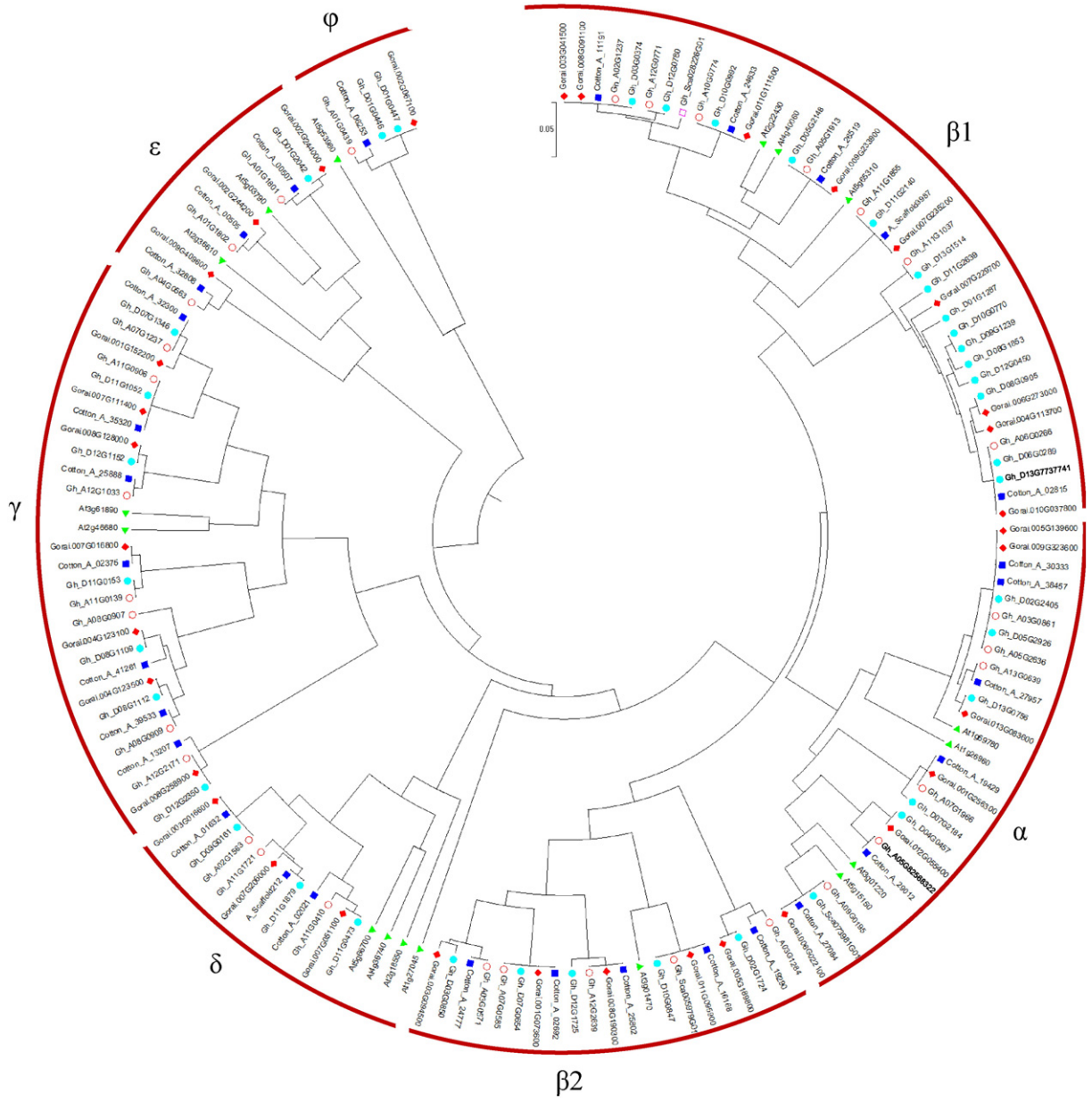

Fig. 1. Phylogenetic tree of the HD-Zip I proteins in *Gossypium*. The tree was generated using protein sequences by the neighbor-joining approach implemented in MEGA6. The clades of α, β1, β2, γ, δ, ε and φ were defined based on the clustering of sequences with known HD-Zip I proteins in Arabidopsis. Genes from Arabidopsis, *G. arboreum*, *G. raimondii*, *G. hirsutum* (A_t) and *G. hirsutum* (D_t) are indicated by green triangles, dark blue squares, red diamonds, red empty circles and light blue solid circles, respectively.

Gh_D12G0780 and *GhHDZ41*/*Gh_D13G7737741*) were newly evolved in *G. hirsutum* (Table 1; Fig. 1). Amongst the HD-Zip I genes found only in *G. hirsutum*, both the A_t and D_t homoeologues of *GhHDZ31* and *GhHDZ36* seemed to have been gained after polyploidization, whereas others were gained only in the D_t subgenome.

Homologues of all the 30 *G. arboreum* HD-Zip I genes were found in the A_t subgenome of *G. hirsutum*, whereas two *G. raimondii* HD-Zip I genes (*Gorai.002G244200* and *Gorai.009G409600*) did not have a homologue in the D_t subgenome of *G. hirsutum*. *Gorai.004G113700* is closely related to *Gorai.006G273000*, and they had only one homologue in *G. hirsutum*. Similarly, the pair of paralogues of *Gorai.008G091100* and *Gorai.003G041500* also had only a single homologue in *G. hirsutum* (Fig. 1; Table 1).

In *G. hirsutum*, HD-Zip I genes were distributed across all 13 pairs of homoeologous chromosomes (Fig. S3). The majority of the A_t and D_t homoeologues were found in corresponding homoeologous chromosomes, while six pairs of the A_t and D_t homoeologues were found in non-homoeologous chromosomes. Five (*GhHDZ6*, *GhHDZ7*, *GhHDZ9*, *GhHDZ10* and *GhHDZ12*) of the six pairs were found between A02/D02 and A03/D03, or between A04/D04 and A05/D05 (Fig. S3), probably due to chromosome translocations observed previously for A02 and A03 as well as A04 and A05 (Rong et al., 2004; Wang et al., 2013).

According to clustering, members of the cotton HD-Zip I subfamily could be divided into seven clades designated α, β1, β2, γ, δ, ε, and φ just as in *Arabidopsis* (Henriksson et al., 2005). To gain insights into the structural diversity of cotton HD-Zip I genes, we analyzed the exon/intron organization of all cotton HD-Zip I genes. We found that homologues of the same gene in the three cotton species usually shared the same gene structure in terms of either the number of exons and introns or the splicing junctions between introns and exons. Only in four cases (*GhHDZ12*, *GhHDZ16*, *GhHDZ18* and *GhHDZ23*), was a discrepancy found (Table S2). For example, of the four homologues of *GhHDZ18*, three (*Gh_A07G1966*, *Gh_D07G2184* and *Gorai.001G256300*) contained three exons and two introns, whereas the *G. arboreum* homologue *Cotton_A_19429* had two exons and one intron. We also found that, in four clades (γ, δ, ε, and φ), the member genes of each individual clade had the same number of exons and introns. For example, all of the clade γ genes had two exons and one intron, and all clade δ genes possessed three exons and two introns (Table S2).

3.3. Evolutionary patterns of the cotton HD-Zip I genes

To gain insights on divergence of the HD-Zip I genes after polyploidization, the nonsynonymous (*dN*) and synonymous (*dS*) nucleotide substitutions and their ratio (*dN*/*dS*) were analyzed for the homologous gene pairs between *G. arboreum* (A_2) and *G. hirsutum* (A_t), as well as those between *G. raimondii* (D_5) and *G. hirsutum* (D_t). Of the 30 pairs of A_2/A_t homologues, three were identical ($dN = dS = 0$), 15 had a $dN/dS < 1$, one (*Cotton_A_26519*/*Gh_A05G1913*) showed $dN = 0$ and $dS > 0$ (two synonymous nucleotide substitutions), and the remaining 11 showed $dN > 0$ and $dS = 0$ (1–5 nonsynonymous nucleotide substitutions). Of the 28 pairs D_5/D_t homologues, one was identical ($dN = dS = 0$), 17 had a $dN/dS < 1$, four showed $dN = 0$ and $dS > 0$ (1–3 synonymous nucleotide substitutions), and four showed $dN > 0$ and $dS = 0$ (2–6 nonsynonymous nucleotide substitutions), the remaining two (*Gorai.009G323600*/*Gh_D05G2926* and *Gorai.008G190300*/*Gh_D12G1725*) had a *dN*/*dS* ratio close to 1 (Table S3). These results suggest that most HD-Zip I genes have evolved mainly under the influence of purifying selection but some (those with $dN > 0$ and $dS = 0$) could have experienced positive selection. It should be noted that not all sequence variations were captured by the analysis of *dN*/*dS*. For example, compared to its *G. raimondii* homologue (*Gorai.002G244000*), *Gh_D01G2042* (*GhOKRA*) had an 8-bp deletion, which caused a frame shift and changed the C-terminal amino acids in *Gh_D01G2042* (Zhu et al,. 2016). In fact, indels were frequently observed in HD-Zip I genes in *G. hirsutum* and *G. arboreum* compared to their homologues in *G. raimondii* (Table S4).

To know whether gene conversion or nonreciprocal homoeologous recombination (NRHR) was involved in evolution of cotton HD-Zip I genes, we did detailed single nucleotide polymorphism (SNP) analysis for the 27 HD-Zip I genes, which have homologues identified in all four (sub)genomes (A_2, D_5, A_t and D_t). We first classified the SNPs identified amongst the four (sub)genomes into two groups (Group I and II) based on whether or not the A_2 and the D_5 genomes have the same nucleotide. Groups I and II SNPs were then further separated into four and seven types, respectively (Fig. 2). Potential A_t-to-D_t conversion (type 10) was found in 17 genes, and potential D_t-to-A_t conversion (type 11) was identified in 22 genes (Table S4). Because a single base type 10 or type 11 change could also be a result of nucleotide mutation, we thus used two consecutive type 10 or type 11 SNPs that are not interrupted by any other intervening mutation to define an NRHR event in this study. Using this criterion, three A_t-to-D_t NRHR events were found in three genes (*GhHDZ8*, *GhHDZ30* and *GhHDZ40*), and six D_t-to-A_t NRHR events were found in five genes (*GhHDZ17*, *GhHDZ21*, *GhHDZ27*, *GhHDZ33* and *GhHDZ37*). *GhHDZ17* contained two NRHR events while other seven genes each contained a single NRHR event. The sequence involved in NRHR ranged from 4 bp (in *GhHDZ37*) to 75 bp (in *GhHDZ17*) with an average length of 32 bp (Table S4).

3.4. Expression patterns of the HD-Zip I genes in Upland cotton

To explore the potential role of HD-Zip I genes in cotton, we analyzed the expression levels of 25 HD-Zip I genes (details in Materials and methods) in roots, cotyledons, −1 to 5 dpa ovules and 15 dpa fibers using qRT-PCR (Fig. 3). Most genes were not expressed or were lowly expressed in roots. More genes were expressed in cotyledons than in roots. *GhHDZ10* and *GhHDZ16* seemed to be predominantly and preferentially expressed in cotyledons, respectively. The majority of genes were expressed in ovules, particularly in −1 to 1 dpa ovules, although the expression levels of about half of the genes were quite low. Of the four genes (*GhHDZ6*, *GhHDZ9*, *GhHDZ27* and *GhHDZ33*) that were relatively highly expressed in ovules, *GhHDZ6* and *GhHDZ27* had their peak expression levels at 3 dpa, while *GhHDZ9* and *GhHDZ33* had their highest expression levels at 1 dpa. It is also worth noting that no gene was preferentially expressed in 15 dpa fibers (Fig. 3). These results suggest that most cotton HD-Zip I genes could play a role in fiber initiation and early fiber development.

To gain more insights into the biological roles of cotton HD-Zip I genes in different tissues, we further analyzed the expression profiles of all *G. hirsutum* HD-Zip I genes using the transcriptome data (PRJNA234454) generated from TM-1. When using FPKM = 8 (bright yellow color in Fig. 4) as the threshold for a gene to be considered as expressed, about half of the 72 *G. hirsutum* HD-Zip I genes were expressed in at least one of the tissues analyzed. Some genes, such as *GhHDZ16* (*Gh_A07G0585*/*Gh_D07G0654*) and *GhHDZ27* (*Gh_A10G0774*/*Gh_D10G0992*), were constitutively expressed in all tissues analyzed, but most showed tissue or developmental stage specificity. For instance, *GhHDZ36* (*Gh_A12G0771*/*Gh_D12G0780*) was mainly expressed in leaf and stem. *GhHDZ14* (*Gh_A05G2636*/*Gh_D05G2926*) was preferentially expressed in petal. *GhHDZ40* (*Gh_A13G0639*/*Gh_D13G0756*) was preferentially expressed in pistil and stamen. *GhHDZ7* (*Gh_A02G1563*/*Gh_D03G0161*) and *GhHDZ32* (*Gh_A11G1721*/*Gh_D11G1879*) were mainly expressed in −1 to 10 dpa ovules. Meanwhile, for the majority genes, their A_t and D_t homoeologues had a similar expression profile in the samples analyzed, although the expression levels of the A_t and D_t homoeologues of several genes, such as *GhHDZ20* (*Gh_A08G0907*/*Gh_D08G1109*) and *GhHDZ31* (*Gh_A11G1037*/*Gh_D13G1514*), were quite different in the tissues analyzed, suggesting a potential functional divergence. Generally, the newly evolved HD-Zip I genes (those not marked on the right side of the gene ID in Fig. 4) in *G. hirsutum* had lower expression levels. In addition, although different cultivars were

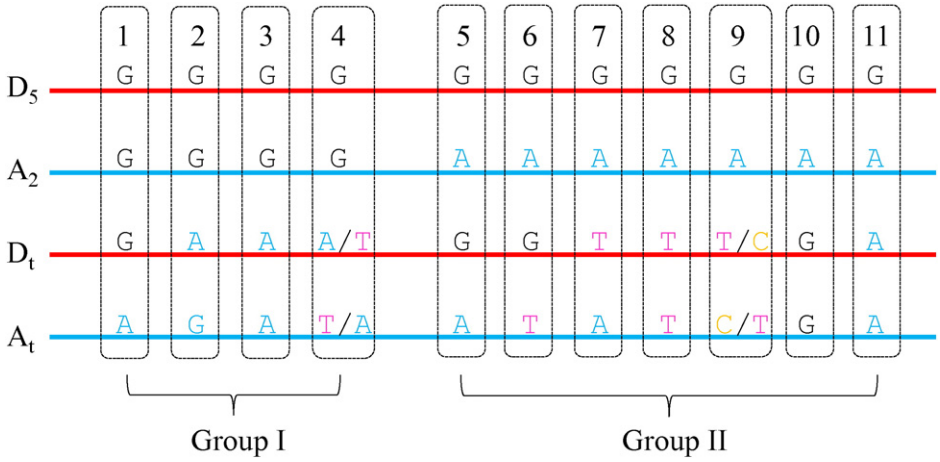

Fig. 2. Schematic diagram showing distinct types of SNPs amongst *G. raimondii*, *G. arboreum* and *G. hirsutum*. SNPs were first separated into two groups. Group I SNPs are those having the same nucleotide in the two diploid progenitors (A_2 and D_5) but with a different nucleotide in at least one of the two subgenomes (A_t and D_t) of the tetraploid. Group II are those having a different nucleotide in the two diploid progenitors (A_2 and D_5) that are maintained or changed in the two subgenomes (A_t and D_t) of the tetraploid. Types 1 and 6: A_t mutated; types 2 and 7: D_t mutated; types 3 and 8: both A_t and D_t mutated but $A_t = D_t$; types 4 and 9: both A_t and D_t mutated but $A_t \neq D_t$; type 5: the genome difference observed in A_2 and D_5 was maintained the two subgenomes (A_t and D_t) of the tetraploid; type 10: potential A_t to D_t conversion; type 11: potential D_t to A_t conversion. Two consecutive type 10 or type 11 conversions that are not interrupted by any other mutation were considered to be a result of a highly confident gene conversion event.

used in the qRT-PCR (MCU-5) and transcriptome (TM-1) analyses, similar expression profiles in roots and ovules were evident for several genes, such as *GhHDZ7* (*Gh_A02G1563*/*Gh_D03G0161*), *GhHDZ8* (*Gh_A03G0571*/*Gh_D03G0850*), *GhHDZ14* (*Gh_A05G2636*/*Gh_D05G2926*), *GhHDZ23* (*Gh_A09G0195*/*Gh_Sca073981G01*) and *GhHDZ32* (*Gh_A11G1721*/*Gh_D11G1879*).

3.5. Cotton HD-Zip I genes in response to *V. dahliae* infection

To investigate whether or not cotton HD-Zip I genes are involved in disease responses, we first analyzed a time course transcriptome dataset (PRJNA234454 downloaded from NCBI; Chen et al., 2015) generated from a *V. dahliae* resistant *G. barbadense* cultivar 7124. The

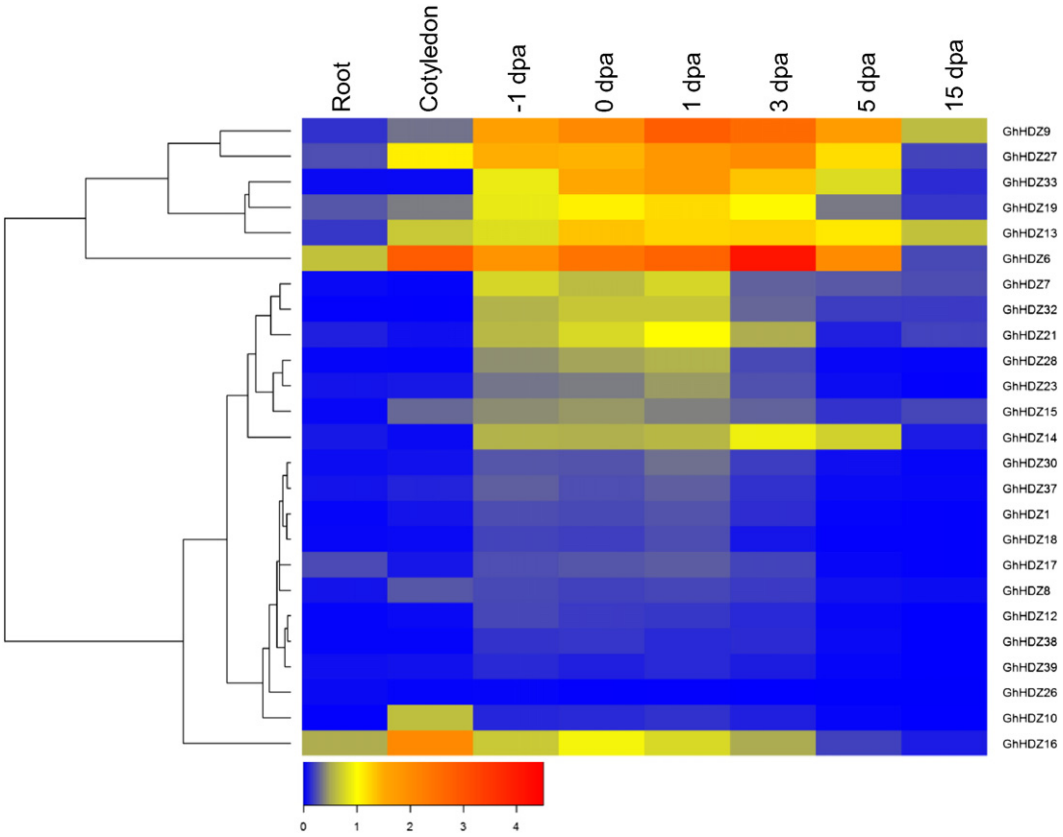

Fig. 3. Expression profiles of the 25 HD-Zip I genes in different tissues from MCU-5. Data shown were average expression levels of three biological replicates each with three technical replicates. The relative expression level of each gene was normalized to that of the cotton ubiquitin gene (GenBank accession no. EU604080) based on the ΔCt approach. Column 1: roots; column 2: cotyledons; columns 3–7 represent ovules at −1, 0, 1, 3 and 5 dpa, respectively; column 8: fibers at 15 dpa.

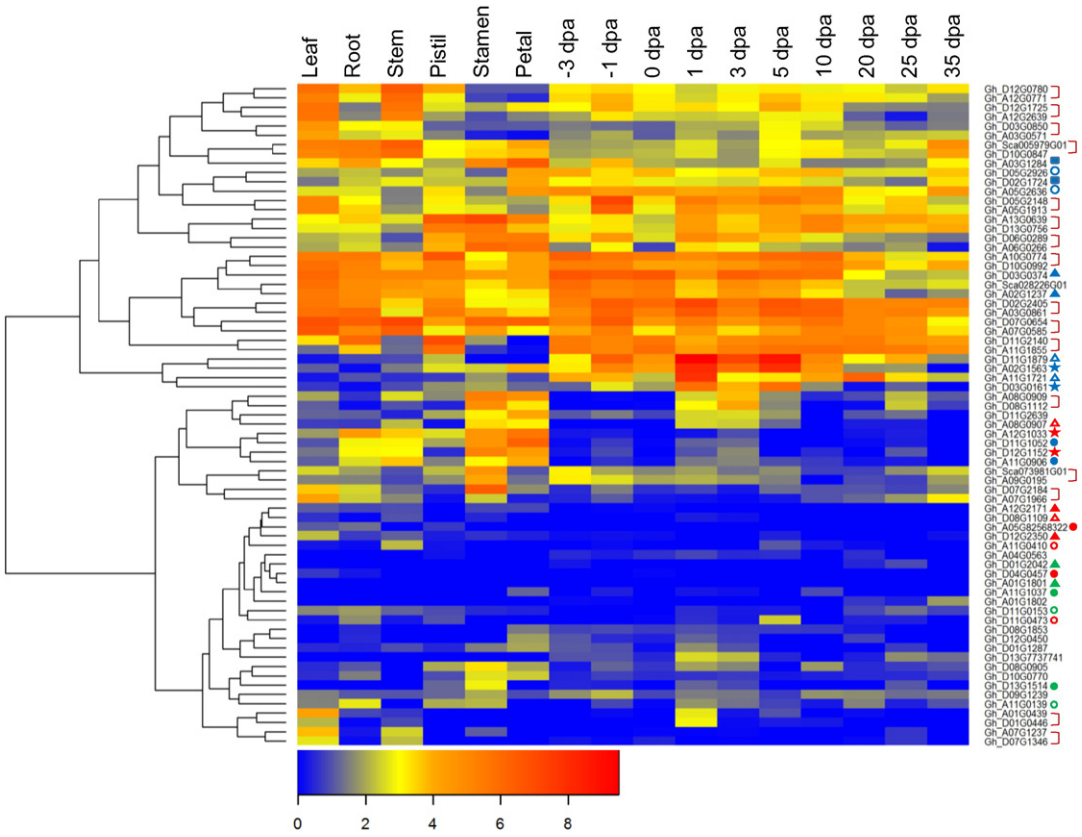

Fig. 4. Expression profiles of the HD-Zip I genes in different tissues of TM-1. Data shown were $\log_2$-transformed FPKM of each gene, which was quantified using the RNA-seq dataset (PRJNA248163) downloaded from NCBI. The A_t and D_t homoeologues of the same gene are indicated by brackets or the same symbols, for instance, a pair of red stars.

Vd-responsive cotton HD-Zip I genes could be classified into five groups. Group I genes were up-regulated by *Vd*-infection at all time points. Group II genes (*GhHDZ1* and *GhHDZ2*) were up-regulated at 2–12 hpi (hours post infection) but down-regulated at 24–72 hpi. Group III genes were down-regulated at 2–12 hpi but up-regulated at 24–72 hpi. Group IV genes (*GhHDZ38*) were constantly down-regulated at all time points. Group V genes were generally down-regulated at all time points upon *Vd*-infection but not as strong as Group IV genes (Fig. 5A). Generally, the A_t and D_t homoeologues of the same gene showed very similar responses following *Vd*-infection.

We then compared the expression changes of 10 genes at 1, 3, and 7 dpi (days post infection) in two *G. hirsutum* cultivars (MCU-5, *Vd* resistant and Siokra 1–4, *Vd* susceptible) by qRT-PCR. After *Vd*-infection, 8 of the 10 genes showed a trend of down-regulation at all three time points in MCU-5 and were up-regulated or unchanged in Siokra 1–4 (Figs. 5B and S4A–F). *GhHDZ7* was unchanged at 1 dpi but up-regulated at 3 and 7 dpi in MCU-5, while up-regulated at all three time points in Siokra 1–4 (Fig. 5B). *GhHDZ21* showed a trend of down-regulation in both MCU-5 and Siokra 1–4 with a more significant down-regulation observed in MCU-5 (Fig. S4G). Comparing the qRT-PCR results of MCU-5 to the RNA-seq results of the *Vd*-resistant *G. barbadense* line, *GhHDZ1*, *GhHDZ9* and *GhHDZ13* showed a consistent response, whereas *GhHDZ6*, *GhHDZ8*, *GhHDZ15*, *GhHDZ21* and *GhHDZ27* showed an opposite response in *Vd*-resistant *G. hirsutum* and *G. barbadense* lines. These results suggest that responses of HD-Zip I genes upon *Vd*-infection could be species and cultivar dependent.

4. Discussion

Our genome-wide survey identified a total of 72 HD-Zip I genes in *G. hirsutum*, 32 and 39 were assigned to the A_t and D_t subgenome, respectively, which is similar to the number of HD-Zip I genes identified in *G. arboreum* (30) and *G. raimondii* (34). All 30 *G. arboreum* and 30 of the 34 *G. raimondii* HD-Zip I genes have their homologues found in the A_t and D_t subgenome of *G. hirsutum*, respectively, suggesting that almost all of the ancestral HD-Zip I genes have been retained in *G. hirsutum* after polyploidization. Two *G. raimondii* genes (*Gorai.004G113700* and *Gorai.008G091100*) that did not have a homologue in *G. hirsutum* are paralogues of *Gorai.006G273000* and *Gorai.003G041500*, respectively, may have been lost after polyploidization. Compared to *G. arboreum* and *G. raimondii*, *G. hirsutum* has two and nine HD-Zip I genes newly evolved in its A_t and D_t subgenome, respectively. The two newly evolved A_t subgenome HD-Zip I genes (*Gh_A11G1037* and *Gh_A12G0771*) also had their D_t homoeologues newly evolved. Although the mechanism for their origin remains uncertain, some of the HD-ZIP I genes newly evolved in the D_t subgenome appear to have been derived by tandem duplication or segmental duplication. For example, *Gh_D01G0446* and *Gh_D01G0447* are identical, which could be a result of tandem duplication (although the possibility of inappropriate genome assembly and/or annotation cannot be ruled out). *Gh_D08G1853*, *Gh_D09G1239*, *Gh_D10G0770*, and *Gh_D12G0450* are very close to each other and all are similar to *Gh_D08G0905* (Fig. 1). They could be a set of paralogues derived from segmental duplication events. In addition, *Gh_D13G7737741* could be a result of segmental duplication of *Gh_D06G0289* (Table 1; Fig. 1). These results suggest that, just as reported for HD-Zip genes in other plant species (Zhao et al., 2011) and other gene families in cotton (Chen et al., 2012, Dou et al., 2014), tandem and segmental duplications could play a substantial role in the expansion of the HD-Zip I gene family during the process of genome evolution. These newly evolved or young genes tend to have a relatively low expression level and to be more tissue specific. For example, *Gh_D10G0770* is mainly expressed in stamens and petals, and *Gh_D12G0450* is mainly expressed in petals (Fig. 4). Whether or not these duplicated genes have different functions in cotton is yet to be investigated. Divergence of expression pattern of *VRS1* and *HvHox2*,

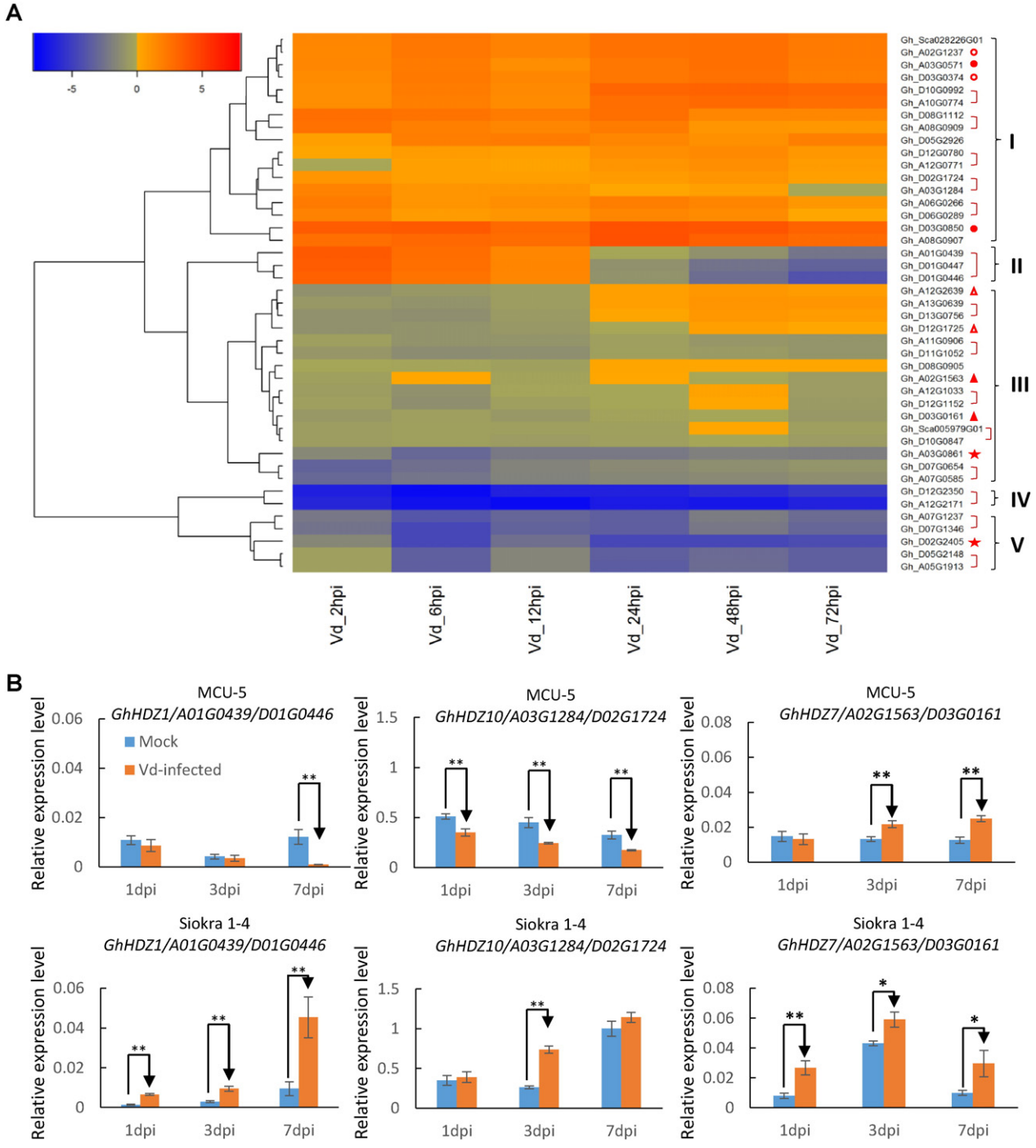

Fig. 5. Expression changes of the HD-Zip I genes in response to *V. dahliae* (*Vd*) infection. (A) Time course expression profiles of the HD-Zip I genes upon *V. dahliae* infection. Data shown were fold changes of FPKM of each gene at each time point (hpi: hours post infection) after *V. dahliae* infection compared to the mock-infected control (2 hpi). FPKM was quantified using the RNA-seq dataset (PRJNA234454) downloaded from NCBI, which was generated from 7124, a *Vd* resistant *G. barbadense* cultivar. The A_t and D_t homoeologues of the same gene are indicated by brackets or the same symbols, for instance, a pair of red stars. (B) Comparison of the expression changes of *GhHDZ9*, *GhHDZ10* and *GhHDZ16* upon *V. dahliae* infection in MCU-5 (*Vd* resistant) and Siokra 1–4 (*Vd* susceptible). Data shown are the relative expression levels based on the average of two biological replicates each with three technical replicates. Error bars represent standard deviations. The relative expression level of each gene was normalized to that of the cotton ubiquitin gene (GenBank accession no. EU604080) based on the ΔCt approach. * and ** indicate a statistically significant difference at $P < 0.05$ and $P < 0.01$, respectively, according to a randomization one-way ANOVA test. dpi: days post infection.

a pair of duplicated HD-Zip I gene in barley, have been shown to contribute to neofunctionalization of *VRS1* (Sakuma et al., 2013).

Gene conversion is a type of nonreciprocal transfer of genetic material in which one segment of DNA contributes genetic information to another, making the recipient location identical to the donor, but not altering the donor sequence (Ohta, 2010). Gene conversion or nonreciprocal homoeologous recombination (NRHR) could be a widespread phenomenon in cotton (Guo et al., 2014). We found that potential NRHR events exist in 25 HD-Zip I genes with eight of them experiencing highly confident NRHR events. The combined effects of gene conversion and point mutation have been proposed to determine the diversification of duplicated genes, with gene conversion playing a major role in accelerating the spread of beneficial mutations through all gene family members (Gjini et al., 2014). In tetraploid cottons, NRHR events were

found to be widespread and biased. A_t-to-D_t NRHR events are more abundant in heterochromatin and are highly correlated with GC content and transposon distribution; by contrast, D_t-to-A_t NRHR events are more abundant in euchromatin and genes, which have been proposed to contribute to the superior yield and fiber quality of tetraploid cottons (Guo et al., 2014). Amongst the 27 HD-Zip I genes, 3 A_t-to-D_t and 6 D_t-to-A_t highly confident NRHR events were observed, consistent with the observation of D_t-to-A_t conversion being more abundant than A_t-to-D_t conversion in genes, although our result was based on a small set of HD-Zip I genes. Our recent work has shown that a D_t-to-A_t NRHR event found in the okra leaf mutant could play a role in the origin of the okra leaf shape in *G. hirsutum* (Zhu et al., 2016).

As a family of plant specific transcription factors, the HD-Zip I genes have been extensively investigated in various plant species and several have been shown to be involved in plant development and stress responses (Henriksson et al., 2005; Harris et al., 2011; Zhao et al., 2011; Hur et al., 2015). In cotton, the expression patterns and functions of a few HD-Zip IV genes have been investigated (Guan et al., 2008; Zhang et al., 2010; Walford et al., 2012; Shan et al., 2014). However, for the HD-Zip I genes, only *Gh_D01G2042* (*GhOKRA*) has recently been demonstrated to be the gene underlying the okra leaf shape in *G. hirsutum* (Zhu et al., 2016), and functions of the remaining HD-Zip I genes are yet to be determined. The expression profiles of the HD-Zip I genes generated in this study should provide clues for their functional characterization. We found that expression of the majority of the HD-Zip I genes could be detected by qPCR and/or transcriptome sequencing in at least one of the tissues analyzed. A few genes seemed to be expressed constitutively in all tissues, but most genes were expressed preferentially in certain tissue(s), such as *GhHDZ36* (*Gh_A12G0771*/*Gh_D12G0780*) mainly in leaf and root, *GhHDZ23* (*Gh_A09G0195*/*Gh_Sca073981G01*) mainly in stamen (Fig. 4). For the two genes that were highly expressed from −3 to 35 dpa ovules, it is *GhHDZ9* (*Gh_A03G0861*/*Gh_D02G2405*) but not *GhHDZ33* (*Gh_A11G1855*/*Gh_D11G2140*) that was also expressed in 15 dpa fibers (Figs. 3 and 4), suggesting that *GhHDZ9* could be involved in fiber initiation and development, and that *GhHDZ33* could be involved in seed development. *GhHDZ9* and *GhHDZ33* belongs to the α and β1 clade, respectively (Fig. 1). *GhHDZ33* is closely related to *ATHB5* and *ATHB6* of *Arabidopsis*. *ATHB5* has been shown to be a positive regulator of seed germination and post-germinative seedling growth by mediating the inhibitory effect of ABA (Johannesson et al., 2003), but is a repressor of the *AUX/IAA* gene *BODENLOS/IAA12* that might contribute to the exclusion of *BODENLOS/IAA12* from epidermis and cortex (De Smet et al., 2013). *ATHB6* is a negative regulator of the ABA signaling pathway because overexpressing *ATHB6* reduces ABA sensitivity in seed germination (Himmelbach et al., 2002). The AUX/IAA proteins interact with AUXIN RESPONSE FACTORs (ARFs) to regulate auxin-responsive genes and higher auxin levels in the ovule epidermis at anthesis promote fiber initials (Zhang et al., 2011a). ABA is known to inhibit fiber initiation in cultured ovules (Zhang et al., 2009), and most likely has to be tightly controlled during fiber initiation (Kim et al., 2015). A role of *GhHDZ9* and *GhHDZ33*, together with *GhHDZ6* and *GhHDZ27* that also belong to the β1 clade and highly expressed in −3 to 10 dpa ovules, in fiber initiation and seed development in the context of auxin signaling as well as ABA metabolism and signaling is worth further exploration.

Expression divergence of homoeologues has been reported in cotton (Adams et al., 2003; Flagel et al., 2008) and has been proposed to play a role in subfunctionalization and neofunctionalization of homoeologues (Flagel et al., 2008; Zhang et al., 2011b). We found that although the expression levels of the A_t and D_t homoeologues of some genes were quite different, such as the expressions of *Gh_A07G1966* and *Gh_D07G2184* (*GhHDZ18*) in leaf, stamen and petal, the expression difference between the A_t and D_t homoeologues of the majority HD-Zip I genes seemed generally to be smaller than that between different genes (Fig. 4). This result suggests that expression divergence of homoeologues in polyploids could be a clue to their possible functional divergence, but a conclusion can only be made by functional characterization of each homoeologue using either natural mutations and/or homoeologue-specific knock-outs. CRISPR/Cas9 based gene editing would be the approach of choice for this purpose because only a 20-bp short sequence is required for target specificity (Belhaj et al., 2015).

Many studies have demonstrated the importance of HD-Zip I genes in responses to abiotic stresses (Henriksson et al., 2005; Harris et al., 2011; Zhao et al., 2011), but little has been done their role in response to biotic stresses (Gao et al., 2014). We investigated expression changes of the cotton HD-Zip I genes following *V. dahliae* infection in both *G. hirsutum* and *G. barbadense*. In the *Vd*-resistant *G. barbadense* cultivar, four types of responses were evident for the HD-Zip I genes (Fig. 5A; details see Results), although most genes showed constant up- or down-regulation upon *Vd*-infection at all time points. In the *Vd*-resistant *G. hirsutum* cultivar, only three (*GhHDZ1*, *GhHDZ9* and *GhHDZ13*) of the 10 analyzed genes showed the same type of response (down-regulation) as that observed in *Vd*-resistant *G. barbadense* cultivar. These three genes were up-regulated or unchanged in the *Vd*-susceptible *G. hirsutum* cultivar. These results suggest that responses of the HD-Zip I genes following *Vd*-infection are species and cultivar dependent, and that the three genes showed a consistent response in *Vd*-resistant *G. hirsutum* and *G. barbadense* should be the good candidates for further investigation.

5. Conclusion

Taking advantage of the recently available genome sequences of three cotton species (*G. hirsutum*, *G. arboreum* and *G. raimondii*), we performed genome-wide analysis of the HD-Zip I genes in cotton, investigated conservation and divergence of cotton HD-Zip I genes and explored the potential biological roles of HD-Zip I genes by analyses of their expression profiles in various tissues and in response to *Vd*-infection. Our results provided a comprehensive view of the cotton HD-Zip I genes and useful information for their further functional characterization.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.plgene.2016.05.002.

Acknowledgments

This study was supported by Cotton Breeding Australia, a joint venture between Cotton Seed Distributors Ltd. and CSIRO, and a scholarship awarded to Dr. Jian Zhang by the China Scholarship Council.

References

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U. S. A. 100, 4649–4654.

Agalou, A., Purwantomo, S., Overnas, E., Johannesson, H., Zhu, X., Estiati, A., de Kam, R.J., Engstrom, P., Slamet-Loedin, I.H., Zhu, Z., Wang, M., Xiong, L., Meijer, A.H., et al., 2008. A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. Plant Mol. Biol. 66, 87–103.

Ariel, F., Diet, A., Verdenaud, M., Gruber, V., Frugier, F., Chan, R., Crespi, M., 2010. Environmental regulation of lateral root emergence in *Medicago truncatula* requires the HD-Zip I transcription factor HB1. Plant Cell 22, 2171–2183.

Ariel, F.D., Manavella, P.A., Dezar, C.A., Chan, R.L., 2007. The true story of the HD-Zip family. Trends Plant Sci. 12, 419–426.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., Nekrasov, V., 2015. Editing plant genomes with CRISPR/Cas9. Curr. Opin. Biotechnol. 32, 76–84.

Cabello, J.V., Arce, A.L., Chan, R.L., 2012. The homologous HD-Zip I transcription factors *HaHB1* and *AtHB13* confer cold tolerance via the induction of pathogenesis-related and glucanase proteins. Plant J. 69, 141–153.

Chen, A., He, S., Li, F., Li, Z., Ding, M., Liu, Q., Rong, J., 2012. Analyses of the sucrose synthase gene family in cotton: structure, phylogeny and expression patterns. BMC Plant Biol. 12, 85.

Chen, X., Chen, Z., Zhao, H., Zhao, Y., Cheng, B., Xiang, Y., 2014. Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. PLoS One 9, e87156.

Chen, Z.J., Scheffler, B.E., Dennis, E., Triplett, B.A., Zhang, T., Guo, W., Chen, X., Stelly, D.M., Rabinowicz, P.D., Town, C.D., Arioli, T., Brubaker, C., Cantrell, R.G., et al., 2007. Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol. 145, 1303–1310.

Chen, J.Y., Huang, J.Q., Li, N.Y., Ma, X.F., Wang, J.L., Liu, C., Liu, Y.F., Liang, Y., Bao, Y.M., Dai, X.F., 2015. Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium wilt*. BMC Plant Biol. 15, 148.

De Smet, I., Lau, S., Ehrismann, J.S., Axiotis, I., Kolb, M., Kientz, M., Weijers, D., Jurgens, G., 2013. Transcriptional repression of *BODENLOS* by HD-ZIP transcription factor HB5 in *Arabidopsis thaliana*. J. Exp. Bot. 64, 3009–3019.

Deng, X., Phillips, J., Meijer, A.H., Salamini, F., Bartels, D., 2002. Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant *Craterostigma plantagineum*. Plant Mol. Biol. 49, 601–610.

Dezar, C.A., Gago, G.M., Gonzalez, D.H., Chan, R.L., 2005. *Hahb-4*, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants. Transgenic Res. 14, 429–440.

Dou, L., Zhang, X., Pang, C., Song, M., Wei, H., Fan, S., Yu, S., 2014. Genome-wide analysis of the *WRKY* gene family in cotton. Mol. Gen. Genomics. 289, 1103–1121.

Flagel, L., Udall, J., Nettleton, D., Wendel, J., 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol. 6, 16.

Frank, W., Phillips, J., Salamini, F., Bartels, D., 1998. Two dehydration-inducible transcripts from the resurrection plant *Craterostigma plantagineum* encode interacting homeodomain-leucine zipper proteins. Plant J. 15, 413–421.

Fryxell, P.A., 1992. A revised taxonomic interpretation of *Gossypium* L. (*Malvaceae*). Rheedea 2, 108–165.

Gao, D., Appiano, M., Huibers, R.P., Chen, X., Loonen, A.E., Visser, R.G., Wolters, A.M., Bai, Y., 2014. Activation tagging of *ATHB13* in *Arabidopsis thaliana* confers broad-spectrum disease resistance. Plant Mol. Biol. 86, 641–653.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wuthrich, K., 1994. Homeodomain-DNA recognition. Cell 78, 211–223.

Gjini, E., Haydon, D.T., David Barry, J., Cobbold, C.A., 2014. Revisiting the diffusion approximation to estimate evolutionary rates of gene family diversification. J. Theor. Biol. 341, 111–122.

Guan, X.Y., Li, Q.J., Shan, C.M., Wang, S., Mao, Y.B., Wang, L.J., Chen, X.Y., 2008. The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the *Arabidopsis GLABRA2*. Physiol. Plant. 134, 174–182.

Guo, H., Wang, X., Gundlach, H., Mayer, K.F., Peterson, D.G., Scheffler, B.E., Chee, P.W., Paterson, A.H., 2014. Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). Genetics 197, 1153–1163.

Harris, J.C., Hrmova, M., Lopato, S., Langridge, P., 2011. Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. New Phytol. 190, 823–837.

Henriksson, E., Olsson, A.S., Johannesson, H., Johansson, H., Hanson, J., Engstrom, P., Soderman, E., 2005. Homeodomain leucine zipper class I genes in *Arabidopsis*, expression patterns and phylogenetic relationships. Plant Physiol. 139, 509–518.

Himmelbach, A., Hoffmann, T., Leube, M., Hohener, B., Grill, E., 2002. Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in *Arabidopsis*. EMBO J. 21, 3029–3038.

Hu, R., Chi, X., Chai, G., Kong, Y., He, G., Wang, X., Shi, D., Zhang, D., Zhou, G., 2012. Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar (*Populus trichocarpa*). PLoS One 7, e31149.

Hur, Y.S., Um, J.H., Kim, S., Kim, K., Park, H.J., Lim, J.S., Kim, W.Y., Jun, S.E., Yoon, E.K., Lim, J., Ohme-Takagi, M., Kim, D., Park, J., et al., 2015. *Arabidopsis thaliana* homeobox 12 (ATHB12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. New Phytol. 205, 316–328.

Johannesson, H., Wang, Y., Engstrom, P., 2001. DNA-binding and dimerization preferences of *Arabidopsis* homeodomain-leucine zipper transcription factors in vitro. Plant Mol. Biol. 45, 63–73.

Johannesson, H., Wang, Y., Hanson, J., Engstrom, P., 2003. The *Arabidopsis thaliana* homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. Plant Mol. Biol. 51, 719–729.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Kim, H.J., Hinchliffe, D.J., Triplett, B.A., Chen, Z.J., Stelly, D.M., Yeater, K.M., Moon, H.S., Gilbert, M.K., Thyssen, G.N., Turley, R.B., Fang, D.D., 2015. Phytohormonal networks promote differentiation of fiber initials on pre-anthesis cotton ovules grown in vitro and in planta. PLoS One 10, e0125046.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., Stein, N., Graner, A., Wicker, T., Tagiri, A., Lundqvist, U., Fujimura, T., Matsuoka, M., et al., 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc. Natl. Acad. Sci. U. S. A. 104, 1424–1429.

Lee, Y.H., Chun, J.Y., 1998. A new homeodomain-leucine zipper gene from *Arabidopsis thaliana* induced by water stress and abscisic acid treatment. Plant Mol. Biol. 37, 377–384.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., et al., 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat. Genet. 46, 567–572.

Liu, W., Fu, R., Li, Q., Li, J., Wang, L., Ren, Z., 2013. Genome-wide identification and expression profile of homeodomain-leucine zipper Class I gene family in *Cucumis sativus*. Gene 531, 279–287.

Manavella, P.A., Arce, A.L., Dezar, C.A., Bitton, F., Renou, J.P., Crespi, M., Chan, R.L., 2006. Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower *Hahb-4* transcription factor. Plant J. 48, 125–137.

Manavella, P.A., Dezar, C.A., Ariel, F.D., Drincovich, M.F., Chan, R.L., 2008. The sunflower HD-Zip transcription factor *HAHB4* is up-regulated in darkness, reducing the transcription of photosynthesis-related genes. J. Exp. Bot. 59, 3143–3155.

Mattsson, J., Ckurshumova, W., Berleth, T., 2003. Auxin signaling in *Arabidopsis* leaf vascular development. Plant Physiol. 131, 1327–1339.

Morelli, G., Ruberti, I., 2000. Shade avoidance responses, driving auxin along lateral routes. Plant Physiol. 122, 621–626.

Mukherjee, K., Burglin, T.R., 2006. MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. Plant Physiol. 140, 1142–1150.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., Takahashi, T., 2006. Characterization of the class IV homeodomain-leucine zipper gene family in *Arabidopsis*. Plant Physiol. 141, 1363–1375.

Ni, Y., Wang, X., Li, D., Wu, Y., Xu, W., Li, X., 2008. Novel cotton homeobox gene and its expression profiling in root development and in response to stresses and phytohormones. Acta Biochim. Biophys. Sin. Shanghai 40, 78–84.

Ohta, T., 2010. Gene conversion and evolution of gene families: an overview. Genes (Basel) 1, 349–356.

Olsson, A.S., Engstrom, P., Soderman, E., 2004. The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in *Arabidopsis*. Plant Mol. Biol. 55, 663–677.

Paterson, A.H., Wendel, J.F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K.C., Shu, S., Udall, J., Yoo, M.J., Byers, R., Chen, W., et al., 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492, 423–427.

Ponting, C.P., Aravind, L., 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. Trends Biochem. Sci. 24, 130–132.

Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N., Clark, S.E., 2005. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. Plant Cell 17, 61–76.

Qin, Y.F., Li, D.D., Wu, Y.J., Liu, Z.H., Zhang, J., Zheng, Y., Li, X.B., 2010. Three cotton homeobox genes are preferentially expressed during early seedling development and in response to phytohormone signaling. Plant Cell Rep. 29, 1147–1156.

Rong, J., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.W., Delmonte, T.A., Ding, X., Garza, J.J., Marler, B.S., Park, C.H., Pierce, G.J., Rainey, K.M., et al., 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166, 389–417.

Rueda, E.C., Dezar, C.A., Gonzalez, D.H., Chan, R.L., 2005. *Hahb-10*, a sunflower homeobox-leucine zipper gene, is regulated by light quality and quantity, and promotes early flowering when expressed in *Arabidopsis*. Plant Cell Physiol. 46, 1954–1963.

Saddic, L.A., Huvermann, B., Bezhani, S., Su, Y., Winter, C.M., Kwon, C.S., Collum, R.P., Wagner, D., 2006. The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. Development 133, 1673–1682.

Sakuma, S., Pourkheirandish, M., Hensel, G., Kumlehn, J., Stein, N., Tagiri, A., Yamaji, N., Ma, J.F., Sassa, H., Koba, T., Komatsuda, T., 2013. Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. New Phytol. 197, 939–948.

Sawa, S., Ohgishi, M., Goda, H., Higuchi, K., Shimada, Y., Yoshida, S., Koshiba, T., 2002. The *HAT2* gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in *Arabidopsis*. Plant J. 32, 1011–1022.

Schena, M., Davis, R.W., 1992. HD-Zip proteins: members of an *Arabidopsis* homeodomain protein superfamily. Proc. Natl. Acad. Sci. U. S. A. 89, 3894–3898.

Schrick, K., Nguyen, D., Karlowski, W.M., Mayer, K.F., 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. Genome Biol. 5, R41.

Sessa, G., Carabelli, M., Sassi, M., Ciolfi, A., Possenti, M., Mittempergher, F., Becker, J., Morelli, G., Ruberti, I., 2005. A dynamic balance between gene activation and repression regulates the shade avoidance response in *Arabidopsis*. Genes Dev. 19, 2811–2815.

Sessa, G., Morelli, G., Ruberti, I., 1993. The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. EMBO J. 12, 3507–3517.

Shan, C.M., Shangguan, X.X., Zhao, B., Zhang, X.F., Chao, L.M., Yang, C.Q., Wang, L.J., Zhu, H.Y., Zeng, Y.D., Guo, W.Z., Zhou, B.L., Hu, G.J., Guan, X.Y., et al., 2014. Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. Nat. Commun. 5, 5519.

Sicard, A., Thamm, A., Marona, C., Lee, Y.W., Wahl, V., Stinchcombe, J.R., Wright, S.I., Kappel, C., Lenhard, M., 2014. Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. Curr. Biol. 24, 1880–1886.

Soderman, E., Hjellstrom, M., Fahleson, J., Engstrom, P., 1999. The HD-Zip gene *ATHB6* in *Arabidopsis* is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions. Plant Mol. Biol. 40, 1073–1083.

Soderman, E., Mattsson, J., Engstrom, P., 1996. The *Arabidopsis* homeobox gene *ATHB-7* is induced by water deficit and by abscisic acid. Plant J. 10, 375–381.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

Tron, A.E., Bertoncini, C.W., Chan, R.L., Gonzalez, D.H., 2002. Redox regulation of plant homeodomain transcription factors. J. Biol. Chem. 277, 34800–34807.

Vlad, D., Kierzkowski, D., Rast, M.I., Vuolo, F., Dello Ioio, R., Galinha, C., Gan, X., Hajheidari, M., Hay, A., Smith, R.S., Huijser, P., Bailey, C.D., Tsiantis, M., 2014. Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. Science 343, 780–783.

Walford, S.A., Wu, Y., Llewellyn, D.J., Dennis, E.S., 2012. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. Plant J. 71, 464–478.

Wang, H., Lin, J., Li, X.G., Chang, Y.H., 2015. Genome-wide identification of pear HD-Zip gene family and expression patterns under stress induced by drought, salinity, and pathogen. Acta Physiol. Plant. 37, 189.

Wang, Y., Henriksson, E., Soderman, E., Henriksson, K.N., Sundberg, E., Engstrom, P., 2003. The *Arabidopsis* homeobox gene, *ATHB16*, regulates leaf development and the sensitivity to photoperiod in *Arabidopsis*. Dev. Biol. 264, 228–239.

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., Paterson, A.H., 2013. A Whole-genome DNA Marker Map for Cotton Based on the D-genome sequence of *Gossypium raimondii* L. G3 (Bethesda). Vol. 3 pp. 1759–1767.

Wendel, J.F., Cronn, R.C., 2003. Polyploidy and the evolutionary history of cotton. Adv. Agron. 78, 139–186.

Zhang, C.H., Ma, R.J., Shen, Z.J., Sun, X., Korir, N.K., Yu, M.L., 2014. Genome-wide analysis of the homeodomain-leucine zipper (HD-ZIP) gene family in peach (*Prunus persica*). Genet. Mol. Res. 13, 2654–2668.

Zhang, F., Zuo, K., Zhang, J., Liu, X., Zhang, L., Sun, X., Tang, K., 2010. An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. J. Exp. Bot. 61, 3599–3613.

Zhang, H.P., Shao, M.Y., Qiao, Z.J., Yuan, S.N., Wang, X.D., Hua, S.J., 2009. Effect of phytohormones on fiber initiation of cotton ovule. Acta Physiol. Plant. 31, 979–986.

Zhang, M., Zheng, X., Song, S., Zeng, Q., Hou, L., Li, D., Zhao, J., Wei, Y., Li, X., Luo, M., Xiao, Y., Luo, X., Zhang, J., et al., 2011a. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. Nat. Biotechnol. 29, 453–458.

Zhang, S., Haider, I., Kohlen, W., Jiang, L., Bouwmeester, H., Meijer, A.H., Schluepmann, H., Liu, C.M., Ouwerkerk, P.B., 2012. Function of the HD-Zip I gene *Oshox22* in ABA-mediated drought and salt tolerances in rice. Plant Mol. Biol. 80, 571–585.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., Hulse-Kemp, A.M., Wan, Q., Liu, B., et al., 2015. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. 33, 531–537.

Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., Magdelenat, G., Couloux, A., Samain, S., Gill, B.S., Rasmussen, J.B., Barbe, V., Faris, J.D., et al., 2011b. Duplication and partitioning in evolution and function of homoeologous *Q* loci governing domestication characters in polyploid wheat. Proc. Natl. Acad. Sci. U. S. A. 108, 18737–18742.

Zhao, Y., Ma, Q., Jin, X., Peng, X., Liu, J., Deng, L., Yan, H., Sheng, L., Jiang, H., Cheng, B., 2014. A novel Maize homeodomain-leucine zipper (HD-Zip) I gene, *Zmhdz10*, positively regulates drought and salt tolerance in both rice and *Arabidopsis*. Plant Cell Physiol. 55, 1142–1156.

Zhao, Y., Zhou, Y., Jiang, H., Li, X., Gan, D., Peng, X., Zhu, S., Cheng, B., 2011. Systematic analysis of sequences and expression patterns of drought-responsive members of the HD-Zip gene family in maize. PLoS One 6, e28488.

Zhu, Q.H., Fan, L., Liu, Y., Xu, H., Llewellyn, D., Wilson, I., 2013. miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. PLoS One 8, e84390.

Zhu, Q.H., Zhang, J., Liu, D., Stiller, W., Liu, D., Zhang, Z., Llewellyn, D., Wilson, I., 2016. Integrated mapping and characterization of the gene underlying the okra leaf trait in *Gossypium hirsutum* L. J. Exp. Bot. 67, 763–774.